Chapter 5
Semantic Technologies for Distributed Search P2P Networks

Amitava Biswas
Texas A&M University, USA

Suneil Mohan
Texas A&M University, USA

Rabi Mahapatra
Texas A&M University, USA

ABSTRACT
Searching is likely to be the next most important service on the Internet after communication. At present centralized Internet search engines can not search a large part of the Internet content, especially those which are behind local search engines. This situation has created disjointed islands of information. P2P networking paradigm has the potential to integrate these under a single unified internet wide search service. However this search infrastructure will have to allow users to perform meaning based search. Therefore the P2P system will need technologies to capture the meaning of what the users intend to search for and then identify relevant objects. This matching between user’s search intention and objects will go beyond simple keyword based comparison. In this chapter the authors present the required techniques to enable a Web architecture that satisfy these needs.

DOI: 10.4018/978-1-61520-686-5.ch005

INTRODUCTION
The Problem of Web Search
Data, information and knowledge are increasingly being made available on the internet. With this information availability, users are ever more resorting to the internet for searching information for everyday use. Around 13 billion search queries are carried out every month and this figure is growing at 38% annually (Patriquin 2009).

With this escalating demand, users’ expectations are also rising. This has been catalyzed by availability of sophisticated search technologies that allow users to pose simple questions like “who designed Titanic ?” to ask.com or provide a simple search
phrases like “designer of Titanic” in Google to get the required information. Users are increasingly expecting search services to cover all possible web sources to precisely retrieve the intended information object (web page or text documents). Instead of a long list of results they prefer a smaller set of more relevant results whenever possible. Users also want to search for non-text objects like pictures, video, audio files, software tools, and web services on the internet. In these cases, they also want to carry out a meaning-based search without having to know the exact description (matching search keywords) of the objects.

This challenge of meaningful and precise searching from a growing mass of heterogeneous collection of objects is an unsolved problem. Today we have large amounts of objects scattered throughout the internet which are partially indexed by general internet search engines and web crawlers like Google and Yahoo (Bergman 2001). In addition, we have multiple local and specialized web search engines, each servicing specific repositories or part of the web, which are not also indexed by general search engines. This fragmentation of the web has created a large number of disjoint islands of information. Users would prefer a unified search service that can search across all these scattered data and services.

**Role of Distributed Search (P2P) Network**

Integration of all these local and specialized search engines and scattered objects will require a search coordination network. The key question is what kind of network topology is most suited for this? A topic-wise hierarchy of search engines and super-search engines (search engines of specialized or local search engines) is not feasible because there is no single hierarchy of topic that can satisfy each and every user and index all the available data. Users will prefer the flexibility of having multiple hierarchies depending on the search being performed. In such situation, an overlay peer-to-peer networking (P2PN) model which is specialized for searching and known as distributed search network (DSN), can be a solution.

**Need for Advanced DSN Mechanism**

However, a simple distributed search network (DSN) with an arbitrary topology alone, is not good enough to solve the problem. The search success rates (recall rates) reported for these DSNs are quite low in order of 5% to 37% (Acosta 2007). Routing a search query from one node to all other nodes in the network that have a matching object may require a large amount of time. In addition, the amount of messages generated within the DSN to percolate this single search query is quite considerable, it is in order of hundreds and thousands of messages per query processed in the DSN (Acosta 2007). To control this end-to-end message routing time and query overhead, additional mechanisms are needed. These mechanisms will deliver optimum message routing and search performances.

**Chapter Theme and Overview**

This chapter focuses on the following: (1) DSN organization mechanisms; and (2) meaning representation techniques that enable meaningful searching using DSNs. To motivate the need for these two, we start by reviewing the basic concepts, performance metrics, and user expectations regarding searching. Next, we narrate the evolution of distributed search network (DSN) designs and identify the missing features for which DSN technologies have not gained mainstream acceptance. Based on this analysis, we illustrate that DSNs need suitable network organization mechanisms and techniques to capture user’s search intention and meaningfully match them with available objects. We do a brief recap on the network science concepts that are relevant to DSN organization. Then we present the design of network organization mechanisms and techniques...
Related Content

Toward A Performing Resource Provisioning Model for Hybrid Cloud
[www.igi-global.com/article/toward-a-performing-resource-provisioning-model-for-hybrid-cloud/210173?camid=4v1a](www.igi-global.com/article/toward-a-performing-resource-provisioning-model-for-hybrid-cloud/210173?camid=4v1a)

Extending Dynamic Scheduling Policies in WorkflowSim by Using Variance based Approach
[www.igi-global.com/article/extending-dynamic-scheduling-policies-in-workflowsim-by-using-variance-based-approach/153971?camid=4v1a](www.igi-global.com/article/extending-dynamic-scheduling-policies-in-workflowsim-by-using-variance-based-approach/153971?camid=4v1a)

Data-Aware Distributed Computing
[www.igi-global.com/chapter/data-aware-distributed-computing/62820?camid=4v1a](www.igi-global.com/chapter/data-aware-distributed-computing/62820?camid=4v1a)

Performance Aware Planning Algorithms for Cloud Environments
[www.igi-global.com/article/performance-aware-planning-algorithms-for-cloud-environments/196264?camid=4v1a](www.igi-global.com/article/performance-aware-planning-algorithms-for-cloud-environments/196264?camid=4v1a)